Paul Over, Alan Smeaton,
and Alen Docef

Eval-Ware: Digital Video Retrieval

As major Web search engines have started to provide video search capabilities as part of their services, it is of particular interest to revisit the topic of video retrieval. In this issue, "Best of the Web" focuses on resources relevant to the evaluation of digital video retrieval algorithms, systems, and components. Such resources for evaluation include frameworks (projects, research programs, standardization activities), data sets (training data, test data, ground truth), evaluation tools, and procedures. (Please send suggestions for Web resources of interest to our readers, proposals for columns, as well as general feedback, by e-mail to Associate Editor Alen Docef at adocef@ vcu.edu.)

THE TREC VIDEO RETRIEVAL EVALUATIONS (TRECVID) FRAMEWORK

www-nlpir.nist.gov/projects/trecvid
[competition-evaluation framework; training/test/ground truth data; evaluation tools, evaluation procedures]
Since 2001, the National Institute of Standards and Technology (see disclaimer) has been running an evaluation of video retrieval systems. This evaluation (known as TRECVID) is an offshoot of the Text Retrieval Conference (TREC) workshop series. Core system tasks being evaluated include shot boundary detection, high-level feature extraction, and automatic and interactive search. In 2002, data was taken from the publicly accessible Internet Movie Archive. In 2003 and 2004, the core data consisted of broadcast news video available only to TRECVID participants via the Linguistic Data Consortium. Some data that was produced by the participants and includ-ed manual annotations of semantic features is publicly available from the "Past Data" area of the TRECVID Web site. Information about evaluation methods, tools, and measures is also available. Tutorials on relevant topics and past evaluation results from participants are available in the TRECVID Workshop archive. [Many TRECVID participants also publish their evaluation studies in the annual ACM Multimedia conference (http://www.sigmm.org/Events/past_acmmm_conferences), the associated Multimedia Information Retrieval workshop (riemann.ist.psu.edu/mir2006), and the Conference on Image and Video Retrieval (www.civr2006.org).]

VIDEO ANALYSIS AND CONTENT EXTRACTION PHASE III PROGRAM

www-nlpir.nist.gov/projects/trecvid/
VACE_PhaseIII-1Feb06_final.pdf
*en.wikipedia.org/wiki/Disruptive_*
Technologies_Office
[research program; data types; require-ments for content extraction]
Video analysis and content extraction (VACE) is a multiyear research program sponsored by the U.S. Intelligence Community and the Defense Department through the Disruptive Technologies Office (DTO) and just beginning the next three-year phase (2006–2009). It aims to continue to advance the state of the art in methods and technologies for automatic video content extraction, intelligent con-tent services, and underlying enabling technologies, in particular those focused on event extraction and understanding. The publicly available broad agency announcement for VACE III includes information about requirements for con-tent extraction, intelligent content servic-es, and enabling technologies with respect to video data types such as broadcast news, surveillance, ground reconnais-sance, and unmanned aerial vehicle video.

THE MPEG-7 STANDARDIZATION FRAMEWORK

www.chiariglione.org/mpeg/standards/m
peg-7/mpeg-7.htm
www.chiariglione.org/mpeg/technolo-
gies/mp07-rsw
isotc.iso.org/livelink/livelink/fetch/2000/2
489/Ittf_Home/PubliclyAvailableStandar
ds.htm
[standardization framework, training/
test data/ground truth data; evaluation
tools; evaluation procedures]
The MPEG-7 standard provides an effi-cient representation for describing multimedia content. Content-based appli-cations such as video search and retrieval are well suited to take advantage of the MPEG-7 representation, which includes visual descriptors for color, texture, shape, motion and localization; and description schemes. The MPEG group has defined descriptors and description schemes as part of the standard and also has coor-dinated the production of a suite of software tools known as the MPEG-7 eXperimentation Model (XM). MPEG-7 data sets that contain news, theater, and other video sequences are available.

VIDEO DATA ARCHIVES

www.archive.org/details/movies
www.open-video.org
www.publicdomaintorrents.com
tvnews.vanderbilt.edu
[archives of video data]
The Internet Movie Archive, mentioned earlier in the context of TRECVID, has gathered a collection of over 30,000 video programs, many contributed under the Creative Commons license (a recent

development that provides some intellectual property protection to content-owners and that is becoming more accepted as a mechanism for distribution of multimedia information). The archive stores content in a variety of video formats, and provides browsing facilities such as thumbnail previewing as well as search through metadata such as title, producer, keywords, etc.

The OpenVideo Project is a publicly available archive of video sequences at the University of North Carolina. It is smaller than the Internet Movie Archive, but has a broader set of search and browse facilities. The archive includes collections from the University of Maryland, Carnegie Mellon University, the 2001 TREC video retrieval test collection, data from the Digital Himalaya project and NASA K-16 science education programs. The repository is intended to be a collection to be used for comparing video retrieval performance. The OpenVideo interface provides video excerpts, video storyboards and fast-forward options through the video.

Another publicly available video repository is the Public Domain Torrents of classic and B-movies. Similarly to the Internet Movie Archive and the OpenVideo project archive, Public Domain Torrents provides a mechanism for text searching through metadata, rating movies, viewing storyboards of keyframes and downloading them in a variety of formats, including those suitable for playback on iPods.

The Vanderbilt University Television News Archive holds more than 30,000 evening news broadcasts from sources such as ABC, CBS, NBC, and CNN. For a fee, this material can be loaned out for research purposes. A valuable resource in the archive is also the abstract that is associated to each story in a regular news program.

VIDEO ANNOTATION TOOLS
FOR SEARCH AND RETRIEVAL
www.alphaworks.ibm.com/tech/ videoannex
www.alphaworks.ibm.com/tech/ multimodalannotation
www.dfki.de/~kipp/anvil
[video annotation tools]
Among the video annotation tools, we include a few examples that are freely available on the Web. The IBM MPEG-7 Annotation Tool and the Multimodal Annotation Tool create MPEG-7 metadata annotations for video sequences. Both tools accept MPEG files as inputs, segment the video sequence into video shots, and allow shots to be annotated with scene descriptions, key object descriptions, event descriptions, and other metadata. Video shot attributes and their annotated descriptions are then saved as MPEG-7 descriptions in an XML file. The Multimodal Annotation Tools is also able to create audio annotations. Both applications are Windows based and available for download with free IBM site registration.

Anvil is a free video annotation tool written in Java that supports hierarchical multilayered annotation driven by user-defined annotation schemes. It is designed for effortless integration with several public domain speech transcription tools. The tool saves annotation information in XML format. After registration, the software is freely available for research and educational purposes.

VIDEO ANALYSIS
AND RETRIEVAL SYSTEMS
www.alphaworks.ibm.com/tech/marvel
www.ee.columbia.edu/dvmm/new Downloads.htm
www.ee.columbia.edu/cuvidsearch
[video retrieval system; sometimes demos only; data sets; software tools]
There exist several video analysis and retrieval systems, each with different characteristics and data sets. Examples include the IBM Alphaworks Multi-media Analysis and Retrieval Engine (MARVEL) from the IBM T.J. Watson Research Center and the online demos from Columbia University. The reader is invited to compare them with other video retrieval systems that are available.

MARVEL is an image and video search tool that indexes, categorizes, and searches large sets of images and videos. It automates the annotation process by using advanced content analysis techniques for labeling and indexing the data. It integrates search methods of multimedia data using feature descriptors, semantic concepts, and metadata. The system is designed to fully support MPEG-7 multimedia content description. The Windows-based application is available for download with free IBM site registration.

Among the online demos from Columbia University's Digital Video Multimedia Lab, the CuVid system is an interactive search engine for broadcast news video that has been developed for TRECVID 2005. It incorporates advanced story segmentation, multimodal retrieval, semantic concept detection, and duplicate detection. Several software tools and data sets are also available for download. The better-known WebSEEK system is a content-based image and video catalog and search tool for the World Wide Web. Its autonomous Web agents collect images and videos and then automatically analyze, index, and assign the images and videos to subject classes. The system has catalogued more than 650,000 images and 10,000 videos from the Web.

AUTHORS
Paul Over (over@nist.gov) is a computer scientist in the National Institute of Standards and Technology's Information Access Division and project leader for the TREC Video Retrieval Evaluations, U.S.A.

Alan Smeaton (alan.smeaton@ computing.dcu.ie) is a professor at Dublin City University and director of the Center for Digital Video Processing, Ireland.

Alen Docef (adocef@vcu.edu) is with the Electrical and Computer Engineering Department at Virginia Commonwealth University.

Disclaimer: Identification of commercial entities, equipment, or materials is not intended to imply recommendation or endorsement by the National Institute of Standards, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose they were presented. **SP**